Group Sequential and Adaptive Clinical Trial Designs:

Achievements and Challenges

Christopher Jennison

Department of Mathematical Sciences,

University of Bath, UK

http://people.bath.ac.uk/mascj

Celebration of Statistics and Data Science

Cornell University

September, 2025

Outline of talk

1. Some reminiscences about Cornell

2. Optimising a group sequential stopping rule

   Problem formulation and optimisation

   Related problems

3. International Conference on Harmonisation: ICH E20, Guideline on Adaptive Designs for Clinical Trials

   General principles

   Inference after a group sequential or adaptive clinical trial

4. Concluding remarks

1. Some reminiscences

As a PhD student at Cornell from 1978 to 1982, I learnt from

Jack Kiefer	Bob Bechhofer	Shayle Searle
Larry Brown	Tom Santner	Walt Federer
Roger Farrell	Lionel Weiss	Doug Robson
Eugene Dynkin	Les Trotter	Philip McCarthy
Bruce Turnbull	Mike Todd	Paul Velleman

Fellow PhD students included

Iain Johnstone	Bob Vanderbei	Walter Piegorsch
Susan Groshen	Radhika Kulkarni	Chuck McCulloch
Costas Gatsonis	Carolyn Lichtenstein	Luke Tierney
Mosuk Chow	Dave Goldsman	Richard Smith

Some reminiscences

In three summer internships at the Dana Farber Cancer Institute in Boston, I met and learnt from

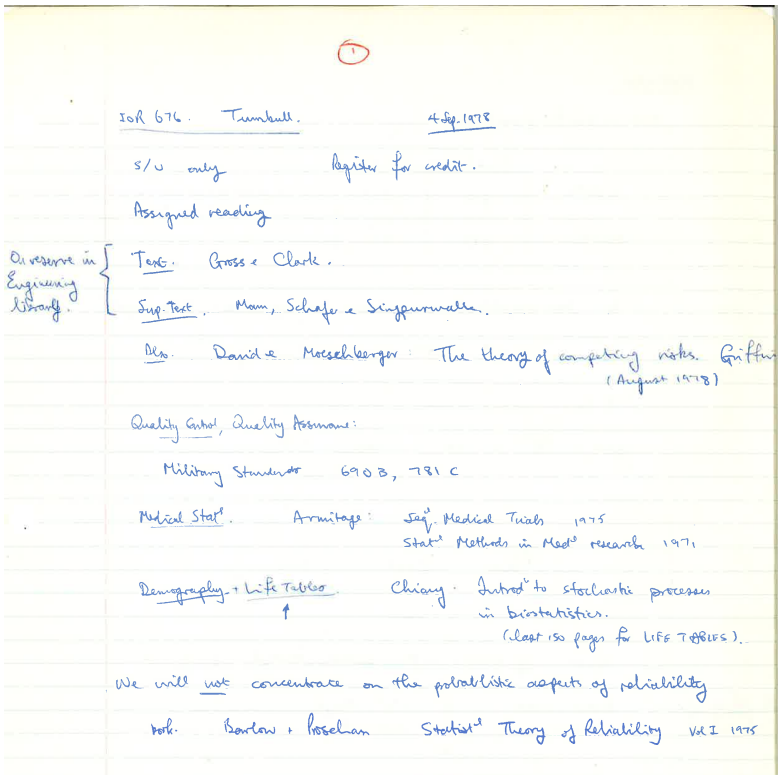
Marvin Zelen	Cyrus Mehta	Janet Barnes
Bill Costello	Nitin Patel	Dave Tritchler
Colin Begg	Dave Schoenfeld	Gregg Dinse
Marcello Pagano	Jim Hanley	Gordon Flowerdew
	Steve Lagakos	

This work taught me about applied data analysis and how clinical trials are conducted.

It helped bridge the gap between academic research and applied statistics.

Some reminiscences

One of the first lectures I attended at Cornell was in the course IOR 676: Survival Data Analysis, given by Bruce Turnbull

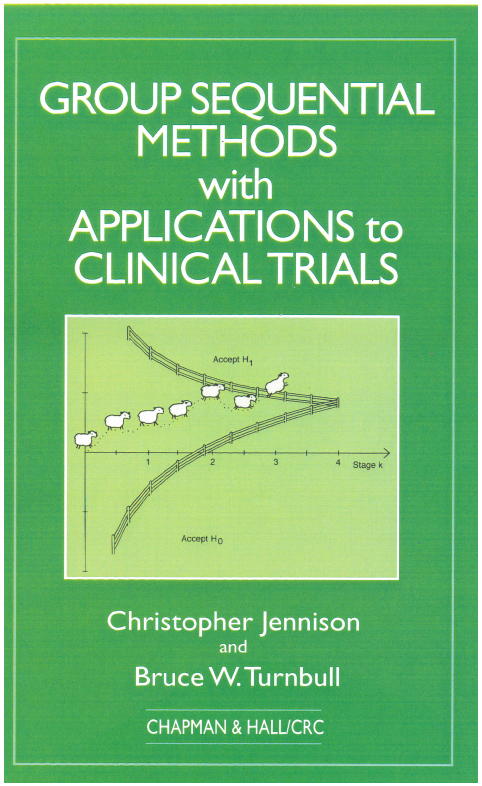


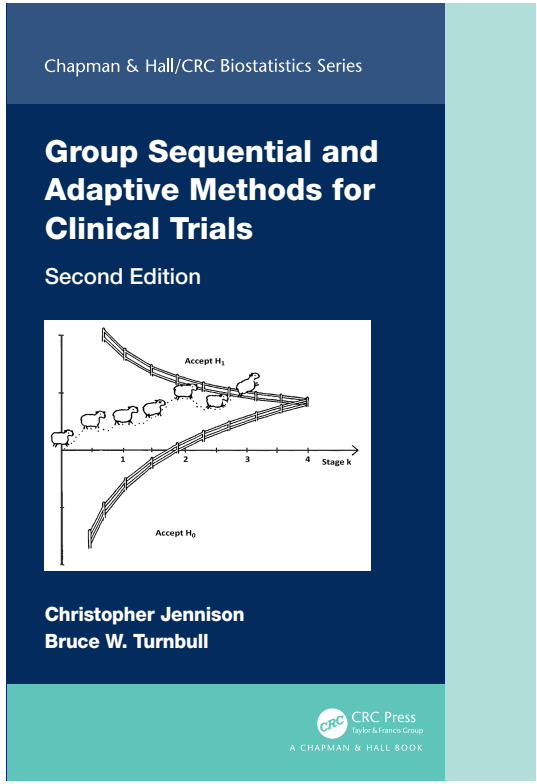
Bruce became my PhD supervisor and we have worked together ever since.

Some reminiscences

A year after I graduated, Bruce Turnbull and Bob Bechhofer invited me back for the summer. This started an annual pattern and I have returned to Ithaca and Cornell almost every year since then.

Bruce and I have written 25 papers together and two books.





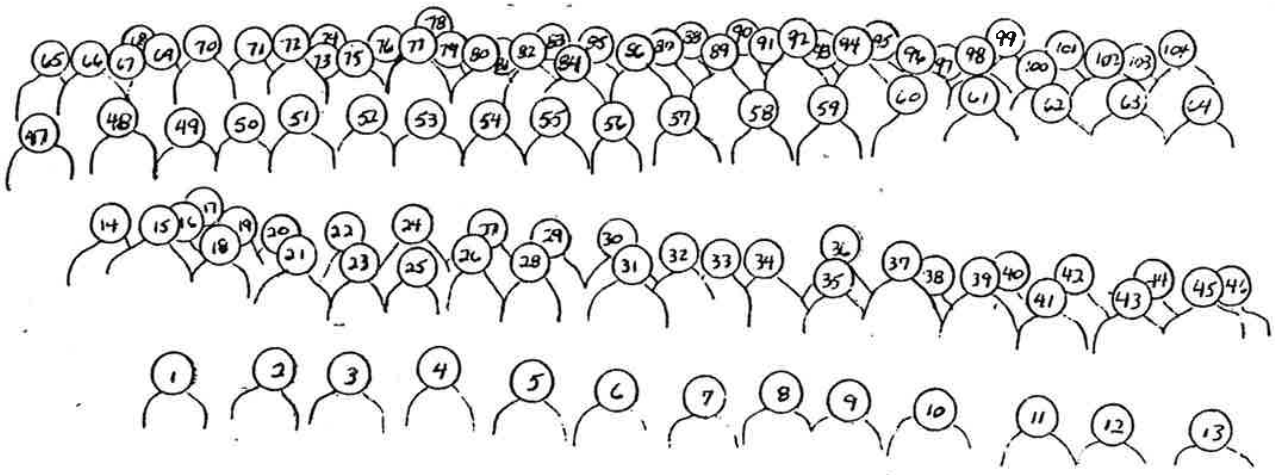
To appear: December 2025

Some reminiscences

What we looked like then — at the Kiefer-Wolfowitz Statistical Research Conference, Cornell, 1983:

JACK KIEFER - JACOB WOLFOWITZ
MEMORIAL
STATISTICAL RESEARCH CONFERENCE

July 6-9, 1983 at Cornell University



(Participants)

1. R. Kulkarni, U. North Carolina
2. E. Seiden, Hebrew U.
3. H.L. Hwang, Northern Illinois
4. J. Hwang, Cornell
5. L. Weiss, Cornell
6. S. Blumenthal, U. Illinois, Urbana
7. H.K. Liu, Cornell
8. P.Y. Chen, Syracuse U.
9. M. Cecce, Cornell
10. C. Gatsonis, Rutgers
11. J.P.M. Schalkwijk, Eindhoven
12. P. Velleman, Cornell
13. R. Adler, Technion
14. C.F. Wu, Wisconsin
15. J. Srivastava, Colorado State
16. R. Smith, Imperial College
17. N. Kiefer, Cornell
18. M. Chow, Cornell
19. W. Sitonik, Cornell
20. G. Legall, Cornell
21. L. Hsu, Santa Barbara
22. W. Piegorsch, Cornell
23. H.F. Wu, Santa Barbara
24. D. Umbach, Cornell
25. B. Taylor, Sherbrooke
26. J. Marden, U. Illinois, Urbana
27. M. Perlman, U. Washington
28. V. Cullinan, Cornell
29. T. Berger, Cornell
30. L. Brown, Cornell
31. A. Cohen, Rutgers
32. W. Strawderman, Rutgers
33. S. Hedayat, U. Illinois, Chicago
34. I. Blumen, Cornell
35. H. Levine, Columbia
36. B. Hajek, U. Illinois, Urbana
37. M. Woodfoofe, #. Michigan
38. S. Searle, Cornell
39. Y.C. Yao, MIT
40. F. Pukelsheim, Freiburg
41. S.P. Lin, Rotkland Research Institute
42. T. Mount, Cornell
43. C. Srinivasan, Kentucky
44. I. Johnstone, Stanford
45. H. Teicher, Rutgers
46. C.C. Heyde, CSIRO
47. G. Casella, Cornell
48. T. Hayter, Cornell
49. A. Tamhane, Northwestern
50. S. Menjoge, Kentucky
51. S. Mishra, U. South Alabama
52. G. Shorrack, Ecole Polytechnique
53. R. Shorrack
54. I. Olkin, Stanford
55. R. Bechhofer, Cornell
56. A. Ash, Boston U.
57. D. Raghavarao, Temple U.
58. W. Federer, Cornell
59. T. Mitchell, Oak Ridge
60. G. Lorden, Cal. Tech.
61. D. Siegmund, Stanford
62. M. Sobel, Santa Barbara
63. W. Studden, Purdue
64.
65. Y. Grize, Cornell
66. P. Huber, Harvard
67. A. Shapiro, Unisa
68. T. Green, Cornell
69. D. Robson, Cornell
70. T.L. Lai, Columbia
71.
72. C. Hagwood, U. Virginia
73. K.F. Yu, Yale
74. A. Ruhkin, Purdue
75. S. Gupta, Purdue
76. D. Whittinghill, Purdue
77. A. Dvoretsky, Hebrew U.
78. T. Cover, Stanford
79. W. Notz, Purdue
80. N. Heckman, SUNY, Stony Brook
81. T. Fine, Cornell
82. S. Schwager, Cornell
83. W.J. Hall, Rochester
84. A. Benjamin, Cornell
85. T. Santner, Cornell
86. B. Turnbull, Cornell
87. K. Mieske, U. Illinois, Chicago
88. J. Sacks, Northwestern
89. C. Jennison, Durham
90. H. Wynn, Imperial College
91. D. Bancroft, Consumer's Union
92. C. McCulloch, Cornell
93. C. Blyth, Queens U.
94. R. Wolpert, Duke
95. J. Bondar, Carleton
96. R. Farrell, Cornell
97. L. Kuo, SUNY, Stony Brook
98. S. Ghosh, U.C., Riverside
99. J. Berger, Purdue
100. C.S. Cheng, Berkeley
101. M. Harel, IUT de Limoges
102. J. Shen, U. Cincinnati
103. G. Constantine, Indiana U.
104. R. Ahlswede, Bielefeld

2. Optimising a group sequential clinical trial

Reference: Jennison & Turnbull, *Kuwait Journal of Science*, 2013.

Consider a Phase 3 clinical trial comparing a new treatment against a standard.

Let θ denote the "effect size", a measure of the improvement in the new treatment over the standard.

We shall test the null hypothesis H_0: $\theta \leq 0$ against $\theta > 0$.

Rejecting H_0 allows us to conclude the new treatment is superior.

We allow type I error probability α for rejecting H_0 when it is true.

We specify power $1 - \beta$ as the probability of rejecting H_0 when $\theta = \delta$. Here δ is, typically, the minimal clinically significant treatment difference.

The trial design, including the method of analysis and stopping rule, must be set up to attain these error rates.

Sequential distribution theory

Let $\widehat{\theta}_k$ denote the estimate of the treatment effect θ at analysis k.

Information for θ at analysis k is $\mathcal{I}_k = \{\text{Var}(\widehat{\theta}_k)\}^{-1},\ k = 1, \ldots, K.$

Canonical joint distribution of $\widehat{\theta}_1, \ldots, \widehat{\theta}_K$

In many situations, in the absence of early stopping, $\widehat{\theta}_1, \ldots, \widehat{\theta}_K$ are approximately multivariate normal,

$$\widehat{\theta}_k \sim N(\theta, \{\mathcal{I}_k\}^{-1}), \quad k = 1, \ldots, K,$$

and

$$\text{Cov}(\widehat{\theta}_{k_1}, \widehat{\theta}_{k_2}) = \text{Var}(\widehat{\theta}_{k_2}) = \{\mathcal{I}_{k_2}\}^{-1} \quad \text{for } k_1 < k_2.$$

References:

Jennison & Turnbull, *JASA*, 1997,

Scharfstein et al, *JASA*, 1997.

An optimal stopping problem

Consider a trial designed to test H_0: $\theta \leq 0$ vs $\theta > 0$, with:

Type I error rate α,

Power $1 - \beta$ at $\theta = \delta$,

Up to K analyses.

A fixed sample test needs information

$$\mathcal{I}_{fix} = \{\Phi^{-1}(\alpha) + \Phi^{-1}(\beta)\}^2/\delta^2.$$

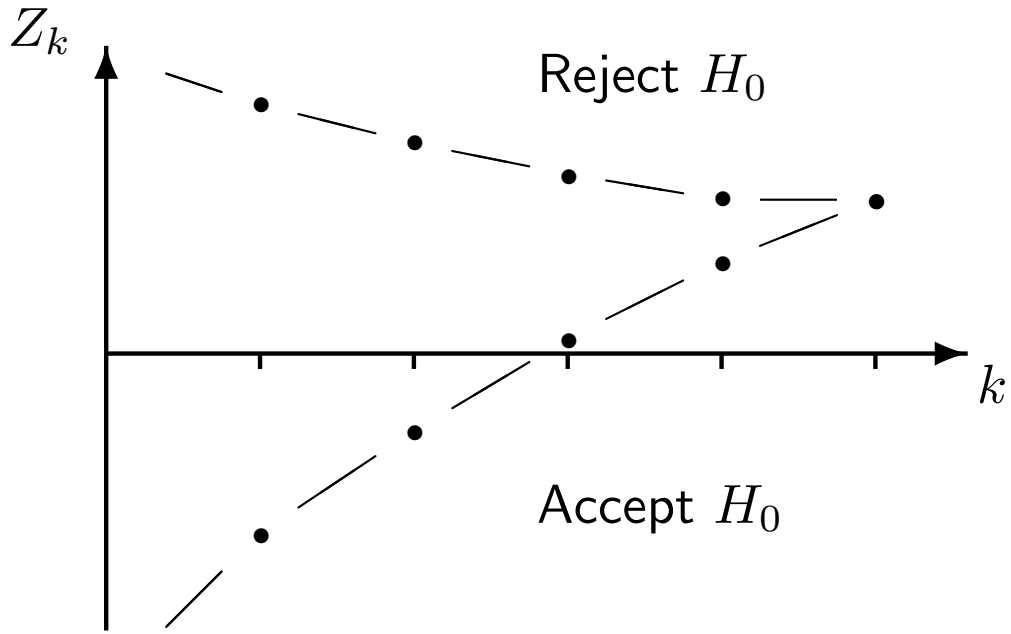
We set the maximum information to be

$$\mathcal{I}_{max} = R\mathcal{I}_{fix},$$

where $R > 1$, with equal increments between analyses.

Optimal group sequential tests

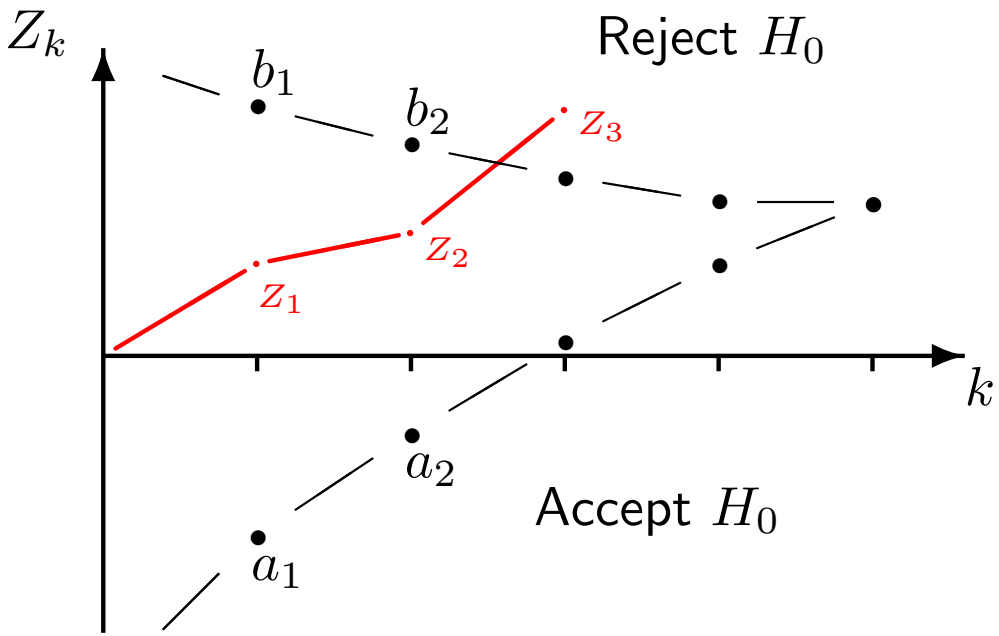
The error rates impose two constraints on the $2K-1$ boundary points — leaving a high dimensional space of possible boundaries.



We shall look for a boundary that minimises

$$\{E_0(\mathcal{I}) + E_\delta(\mathcal{I})\}/2.$$

Computations for group sequential tests



We need to be able to calculate the probabilities of basic events such as

$$a_1 < Z_1 < b_1, \quad a_2 < Z_2 < b_2, \quad Z_3 > b_3.$$

Combining such probabilities gives key properties, such as $Pr_\theta\{\text{Reject } H_0\}$ and $E_\theta(\mathcal{I})$.

Numerical integration

We can write probabilities as nested integrals, e.g.,

$$Pr\{a_1 < Z_1 < b_1,\, a_2 < Z_2 < b_2,\, Z_3 > b_3\} =$$

$$\int_{a_1}^{b_1} \int_{a_2}^{b_2} \int_{b_3}^{\infty} f_1(z_1)\, f_2(z_2|z_1)\, f_3(z_3|z_2)\, dz_3\, dz_2\, dz_1.$$

Applying numerical integration, we replace each integral by a sum of the form

$$\int_a^b f(z)\, dz \;=\; \sum_{i=1}^{n} w(i)\, f(z(i)),$$

where $z(1), \ldots, z(n)$ is a grid of points from a to b.

Numerical integration

Thus, we have

$$Pr\{a_1 < Z_1 < b_1,\, a_2 < Z_2 < b_2,\, Z_3 > b_3\} \approx$$

$$\sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \sum_{i_3=1}^{n_3} w_1(i_1)\, f_1(z_1(i_1))\; w_2(i_2) f_2(z_2(i_2)|z_1(i_1))$$

$$w_3(i_3)\; f_3(z_3(i_3)|z_2(i_2)).$$

Multiple integrations and summations will arise, e.g., for an outcome at analysis k,

$$\sum_{i_1=1}^{n_1} \ldots \sum_{i_k=1}^{n_k} w_1(i_1)\, f_1(z_1(i_1))\; w_2(i_2) f_2(z_2(i_2)|z_1(i_1))$$

$$\ldots\; w_k(i_k)\; f_k(z_k(i_k)|z_{k-1}(i_{k-1})).$$

Numerical integration

In the multiple summation

$$\sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}\cdots\sum_{i_k=1}^{n_k} w_1(i_1)\, f_1(z_1(i_1))\; w_2(i_2) f_2(z_2(i_2)|z_1(i_1))$$

$$\cdots\; w_k(i_k)\; f_k(z_k(i_k)|z_{k-1}(i_{k-1})),$$

the structure of the k nested summations is such that the computation required is of the order of $k-1$ double summations.

Using Simpson's rule with 100 to 200 grid points per integral can give accuracy to 5 or 6 decimal places.

For details of efficient sets of grid points, see Ch. 19 of *Group Sequential Methods with Applications to Clinical Trials* by Jennison and Turnbull (2000).

Finding optimal group sequential tests

Recall, we want a group sequential test of H_0: $\theta \leq 0$ vs $\theta > 0$ with

$$Pr_{\theta=0}\{\text{Reject } H_0\} = \alpha,$$

$$Pr_{\theta=\delta}\{\text{Accept } H_0\} = \beta,$$

Analyses at $\mathcal{I}_k = (k/K)\,\mathcal{I}_{max}, \quad k = 1, \ldots, K,$

Minimum possible value of $\{E_0(\mathcal{I}) + E_\delta(\mathcal{I})\}/2.$

We deal with constraints on error rates by introducing Lagrangian multipliers to create the *unconstrained problem* of minimising

$$\{E_0(\mathcal{I}) + E_\delta(\mathcal{I})\}/2 + \lambda_1 Pr_{\theta=0}\{\text{Reject } H_0\} + \lambda_2\, Pr_{\theta=\delta}\{\text{Accept } H_0\}.$$

We shall find a pair of multipliers (λ_1, λ_2) such that the solution has type I and II error rates α and β, then this design will solve the *constrained problem* too.

Bayesian interpretation of the Lagrangian approach

Suppose we put a prior on θ with $Pr\{\theta = 0\} = Pr\{\theta = \delta\} = 0.5$ and specify costs of

$$
\begin{array}{ll}
1 & \text{per unit of information observed,} \\
2\,\lambda_1 & \text{for rejecting } H_0 \text{ when } \theta = 0, \\
2\,\lambda_2 & \text{for accepting } H_0 \text{ when } \theta = \delta.
\end{array}
$$

Then, the total Bayes risk is

$$\{E_0(\mathcal{I}) + E_\delta(\mathcal{I})\}/2 + \lambda_1\, Pr_{\theta=0}\{\text{Reject } H_0\} + \lambda_2\, Pr_{\theta=\delta}\{\text{Accept } H_0\},$$

just as in the Lagrangian problem.

An advantage of the Bayes interpretation is that it can give insight into solving the problem by using "Dynamic Programming" or "Backwards Induction".

Solution by Dynamic Programming

Denote the posterior distribution of θ given $Z_k = z_k$ at analysis k by

$$p^{(k)}(\theta|z_k), \quad \theta = 0, \delta.$$

At the final analysis, K

There is no further sampling cost, so compare decisions

$$\text{Reject } H_0: \qquad E(\text{Cost}) = 2\,\lambda_1\,p^{(K)}(0|z_K),$$

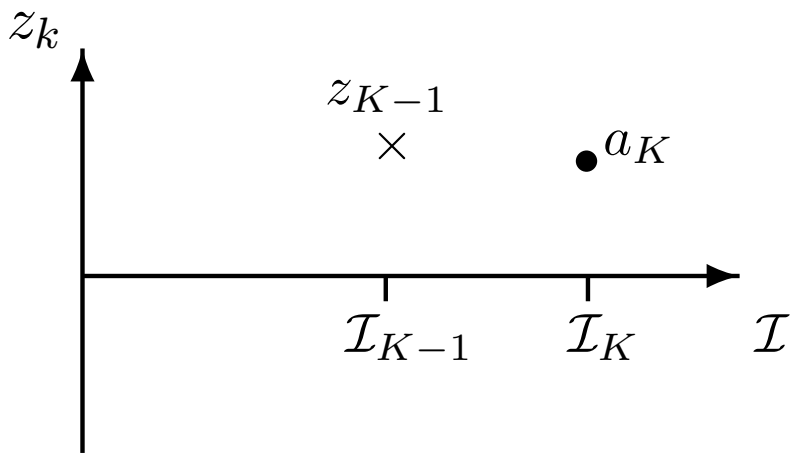
$$\text{Accept } H_0: \qquad E(\text{Cost}) = 2\,\lambda_2\,p^{(K)}(\delta|z_K).$$

The boundary point a_K is the value of z_K where these expected losses are equal.

The optimum decision rule is to reject H_0 for $Z_K > a_K$.

Dynamic Programming

At analysis $K-1$



If the trial stops at this analysis, there is no further cost of sampling and the expected additional cost is

Reject H_0: $\quad 2\,\lambda_1\,p^{(K-1)}(0|z_{K-1})$,

Accept H_0: $\quad 2\,\lambda_2\,p^{(K-1)}(\delta|z_{K-1})$.

At analysis $K-1$

If the trial continues to analysis K, the expected additional cost is

$$1 \times (\mathcal{I}_K - \mathcal{I}_{K-1})$$

$$+ 2\,\lambda_1\, p^{(K-1)}(0|z_{K-1})\, Pr_{\theta=0}\{Z_K > a_K | Z_{K-1} = z_{K-1}\}$$

$$+ 2\,\lambda_2\, p^{(K-1)}(\delta|z_{K-1})\, Pr_{\theta=\delta}\{Z_K < a_K | Z_{K-1} = z_{K-1}\}.$$

We can now define the optimal boundary points:

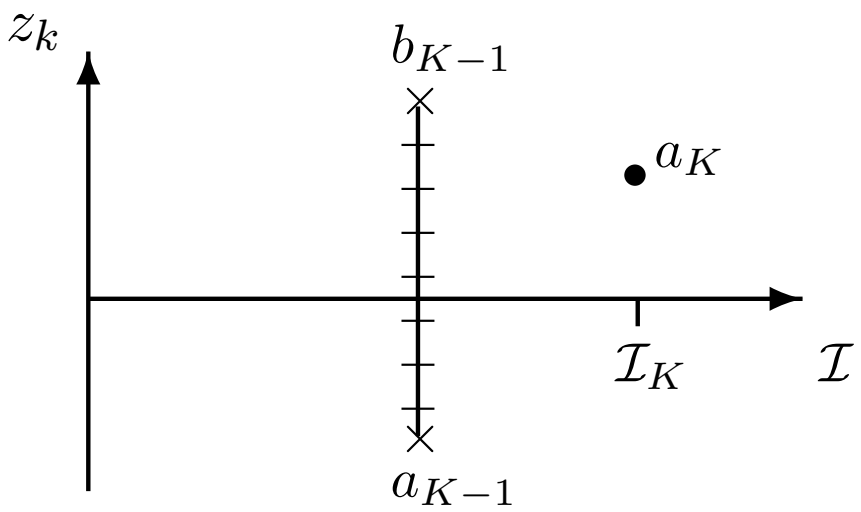
Set b_{K-1} to be the value of z_{K-1} where

$E(\text{Cost of continuing}) \;=\; E(\text{Cost of stopping to reject } H_0).$

Set a_{K-1} to be the value of z_{K-1} where

$E(\text{Cost of continuing}) \;=\; E(\text{Cost of stopping to accept } H_0).$

At analysis $K-1$



Before leaving analysis $K-1$, we set up a grid of points for use in numerical integration over the range a_{K-1} to b_{K-1}.

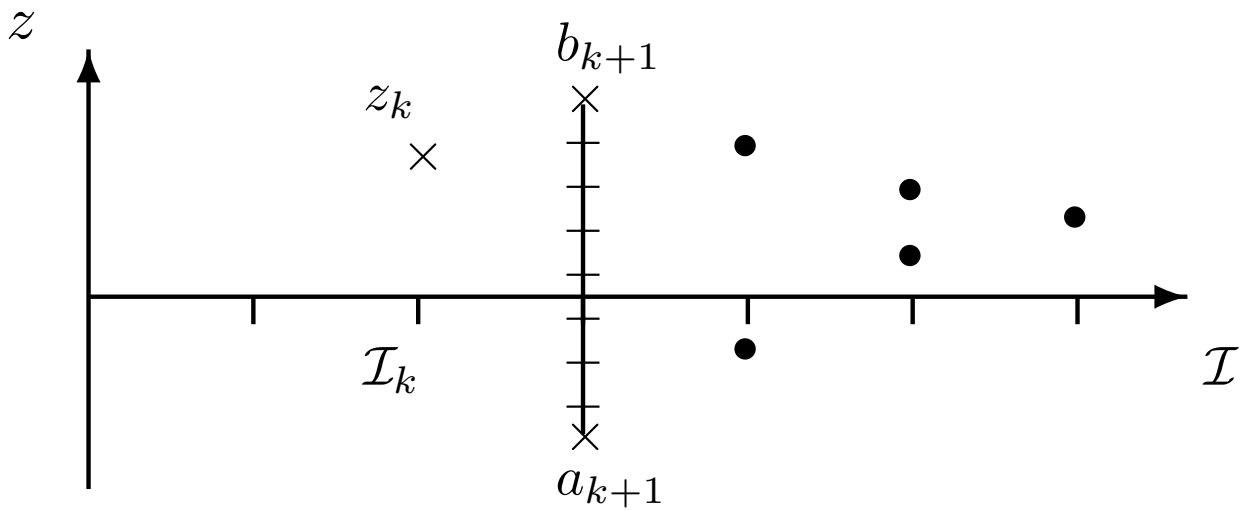
For each point, we sum over the posterior distribution of θ to calculate

$$\beta^{(K-1)}(z_{K-1}) = E(\text{Additional cost when continuing} \,|\, Z_{K-1} = z_{K-1}).$$

We are now ready to move back to analysis $K-2$.

Analyses 1 to $K-2$

We work back through analyses $k = K-2,\ K-3,\ \ldots,\ 1$.



At each analysis, we find the optimal stopping boundary using knowledge of the optimal stopping rule at future analyses.

Then, for a grid of values of z_k, compute

$$\beta^{(k)}(z_k) = E(\text{Additional cost when continuing} \,|\, Z_k = z_k)$$

to use in evaluating the option of continuing at analysis $k-1$.

Solving the original problem

For any given (λ_1, λ_2) we can find the Bayes optimal design and compute its type I and II error rates.

We now search for a pair (λ_1, λ_2) for which type I and type II error rates of the optimal design equal α and β, respectively.

The resulting design will be the optimal group sequential test, with the specified frequentist error rates, for our original problem.

Notes

1. The method of solving the overall problem demonstrates explicitly that good frequentist procedures should be similar to Bayes procedures.

2. The prior and costs in the final Bayes problem are a means to an end, rather than "true" costs of type I and type II errors, or costs of treating patients in the trial.

Properties of optimal designs

Tests with $\alpha = 0.025$, $1 - \beta = 0.9$, K analyses, $\mathcal{I}_{max} = R\mathcal{I}_{fix}$, and equal group sizes, that minimise $\{E_0(\mathcal{I}) + E_\delta(\mathcal{I})\}/2$.

Minimum values of $\{E_0(\mathcal{I}) + E_\delta(\mathcal{I})\}/2$, as a percentage of $\mathcal{I}_{fix}$

	R					
K	1.01	1.05	1.1	1.2	1.3	*Minimum over R*
2	80.8	74.7	73.2	73.7	75.8	73.0 at R=1.13
5	72.2	65.2	62.2	59.8	59.0	58.8 at R=1.38
10	69.2	62.2	59.0	56.3	55.1	54.2 at R=1.6
20	67.8	60.6	57.5	54.6	53.3	51.7 at R=1.8

Observe: $E(\mathcal{I}) \searrow$ as $K \nearrow$ but with diminishing returns,

$E(\mathcal{I}) \searrow$ as $R \nearrow$ up to a point.

Optimisation problems we have addressed

One-sided and two-sided group sequential tests

Eales & Jennison, *Biometrika*, 1992

Eales & Jennison, *Sequential Analysis*, 1995

Barber & Jennison, *Biometrika*, 2002

Group sequential tests with data dependent group sizes

Jennison & Turnbull, *Biometrika*, 2006

Group sequential tests of superiority and non-inferiority

Öhrn & Jennison, *Statistics in Medicine*, 2010

Group sequential tests for delayed responses

Hampson & Jennison, *JRSS, B*, 2013

Optimising gain functions from financial models

Robbie Peck, *University of Bath, PhD thesis*, 2020

Optimal frequentist and Bayes designs coincide

We have seen that computationally, we can find an optimal frequentist group sequential design by solving a Bayes sequential decision problem.

This demonstrates the results of Complete Class Theorems which state that

The class of admissible frequentist designs $=$

The class of Bayes optimal designs

See Brown, Cohen & Strawderman (*Annals of Statistics*, 1980).

Following a Bayes approach but calibrating the problem so the resulting procedure has specified type I and type II error probabilities should lead to the optimal frequentist group sequential design with these error probabilities that minimises the same function of expected sample size.

3. ICH E20: Adaptive Designs for Clinical Trials

The International Conference on Harmonisation (ICH) brings together statisticians from regulatory bodies and pharmaceutical companies to develop guidelines for the drug development process.

The draft ICH E20 Guideline on Adaptive Designs for Clinical Trials was published in June 2025. It lays down general principles, rather than specifying particular designs.

Examples of adaptive designs include

- Group sequential tests stopping for efficacy or futility,
- Adaptive trials with sample size re-assessment,
- Adaptive trials testing multiple hypotheses:
  - Seamless Phase 2-3 trials with treatment selection,
  - Multi-arm multi-stage (MAMS) designs,
  - Enrichment designs.

ICH E20: General principles

The Guideline defines five principles.

1. Adequacy within the development program:

   Justifying the selected dose, etc.

2. Adequacy of trial planning:

   Pre-planned, as simple as possible, some flexibility.

3. Limiting the chances of erroneous conclusions:

   Type I error control.

4. Reliability of estimation:

   Estimates and confidence intervals for cost-benefit decisions.

5. Maintenance of trial integrity:

   Blinding, avoiding information leakage, role of IDMCs.

Inference on termination

Methods are available for some, but not all, cases.

Type of design	Treatment effect estimate with negligible bias	Confidence interval, median unbiased estimate
Group sequential test (GST)	✓	✓
GST + sample size re-assessment (fixed)	✓	✓
GST + sample size re-assessment (flexible)	$\sim$ ✓	$\sim$ ✓
Phase 2/3 with treatment selection, always selecting 1 treatment for Phase 3	✓	✓
Phase 2/3 with treatment selection, selecting 1 or more treatments for Phase 3	✓	✓?
Multi-arm multi-stage design (fixed)	✓	✓?
Multi-arm multi-stage design (flexible)	?	?
Enrichment design (flexible)	✓	✓?

Emerging topics: Inference on termination

Two common difficulties arise.

(i) Flexible designs

Since the sample space is not completely known, frequentist properties, such as the expected value of an estimate, cannot be calculated.

(ii) Confidence intervals in multiple testing problems

A confidence interval for a parameter is the result of testing a family of hypotheses concerning all possible parameter values.

A multiple testing procedure may "exhaust" the type I error probability α, leaving nothing to test non-null parameter values.

This can lead to the situation where H_0: $\theta \leq 0$ is rejected but the upper confidence interval for θ is $(0, \infty)$ — an uninformative confidence interval

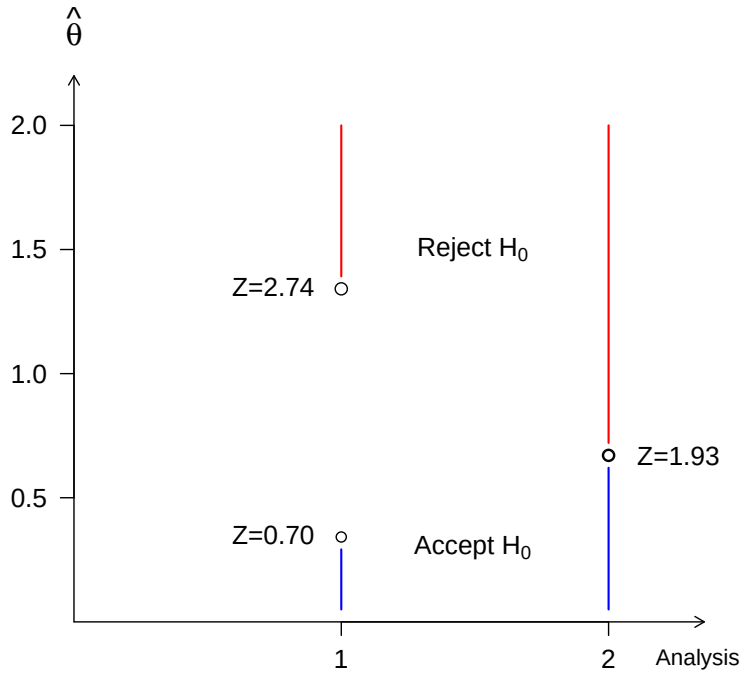
Emerging topics: Inference on termination

Point estimates

ICH E20 refers to estimates that are unbiased or have small bias.

While being exactly unbiased may seem desirable, estimators that achieve this can have some strange properties.

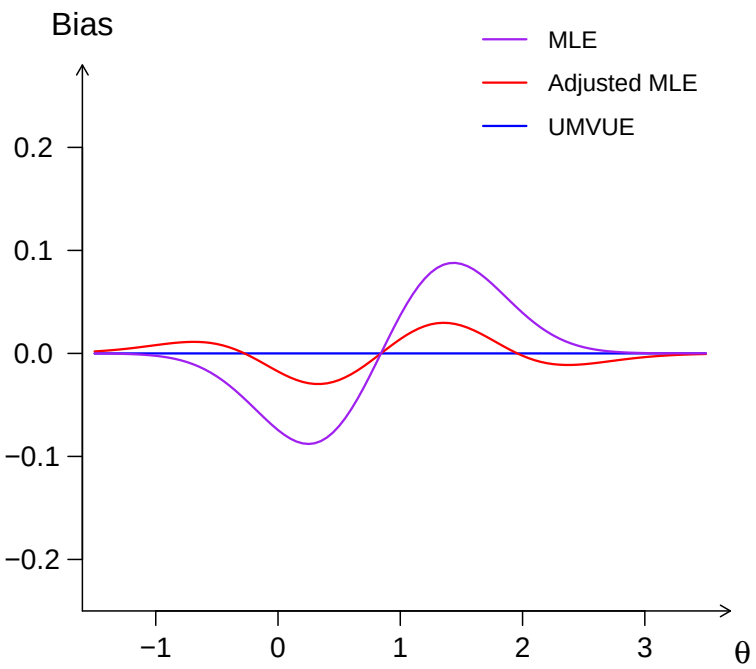
Consider a group sequential test with two analyses.



The test of H_0: $\theta \leq 0$ against $\theta > 0$ has power 0.8 for $\theta = 1$.

Emerging topics: Point estimates on termination

The bias of several estimates:



We denote the maximum likelihood estimate (MLE) on termination by $\widehat{\theta}$.

The Adjusted MLE is formed by subtracting the bias when $\theta = \widehat{\theta}$ from $\widehat{\theta}$.

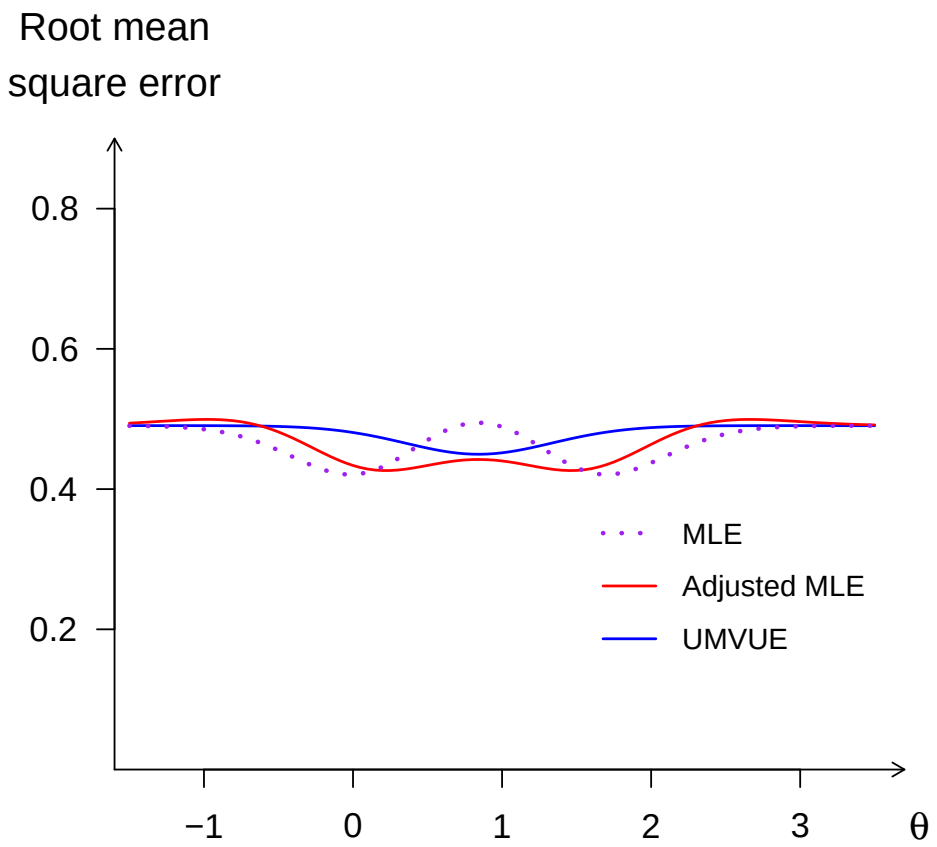
(Whitehead, *Bmka*, 1986)

The Uniform Minimum Variance Unbiased Estimate (UMVUE) uses the fact that $\widehat{\theta}_1$, the MLE at analysis 1, is unbiased for θ.

Applying "Rao-Blackwellization", the UMVUE is the conditional expectation of $\widehat{\theta}_1$, given the final data.

Emerging topics: Point estimates on termination

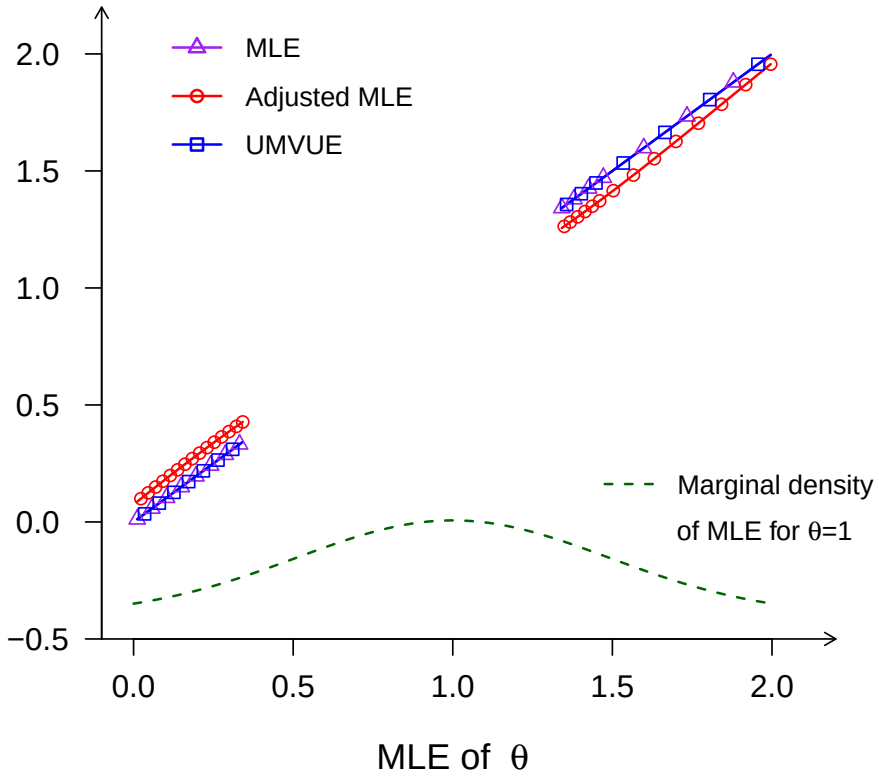
The root mean square error of several estimates:



The UMVUE has a higher variance than the Adjusted MLE, and this results in a higher mean square error.

Emerging topics: Point estimates on termination

Estimates on termination at analysis 1

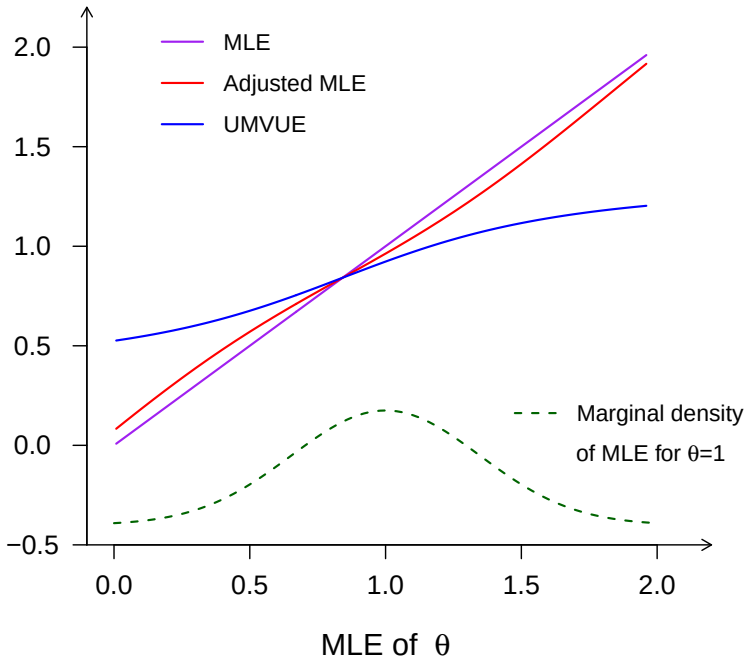


When stopping at analysis 1, the UMVUE is $\widehat{\theta}_1$, the MLE of θ.

So there is no "adjustment for bias" in the UMVUE!

Emerging topics: Point estimates on termination

Estimates on termination at analysis 2



When stopping at analysis 2, the UMVUE can be substantially lower than the MLE.

If the MLE is $\widehat{\theta}_2 = 1.5$, the UMVUE is only 1.12 — but the bias in the MLE is at most 0.09 for any value of θ.

Emerging topics: Point estimates on termination

In more complex adaptive designs, bias may arise

(i) from selecting a treatment arm or patient sub-population based on promising early results,

(ii) from early stopping on a "random high".

Some of the methods proposed for estimation after such trials also use Rao-Blackwellization to find UMVUE or Uniform Minimum Variance Conditionally Unbiased Estimates.

Given the behaviour of the UMVUE estimate in our simple example of a two-stage group single trial with a single parameter to estimate, we should look more closely at how these estimates may behave.

Adjusted estimates with a small bias may well be preferable.

4. Concluding remarks: Thank you, Cornell

I benefitted tremendously from the training I received as a Cornell student, many subsequent research visits, and my long and productive collaboration with Bruce Turnbull.

In welcoming students from around the world, Cornell promotes learning, instigates path-breaking research, and nurtures international collaboration.

In return, Cornell reaps the rewards of a vibrant and dynamic research community.

Today's environment is not an easy one.

I trust Cornell will ride the current wave of adversity and prosper again under a more enlightened administration.